A Toolkit for Multimodal Interface Design: An Empirical Investigation

Dimitrios Rigas and Mohammad Alsuraihi

School of Informatics, University of Bradford, Richmond Road, Bradford, UK
d.rigas@bradford.ac.uk, mmnalsur@bradford.ac.uk

Abstract. This paper introduces a comparative multi-group study carried out to investigate the use of multimodal interaction metaphors (visual, oral, and aural) for improving learnability (or usability from first time use) of interface-design environments. An initial survey was used for taking views about the effectiveness and satisfaction of employing speech and speech-recognition for solving some of the common usability problems. Then, the investigation was done empirically by testing the usability parameters: efficiency, effectiveness, and satisfaction of three design-toolkits (TVOID, OFVOID, and MMID) built especially for the study. TVOID and OFVOID interacted with the user visually only using typical and time-saving interaction metaphors. The third environment MMID added another modality through vocal and aural interaction. The results showed that the use of vocal commands and the mouse concurrently for completing tasks from first time use was more efficient and more effective than the use of visual-only interaction metaphors.

Keywords: interface-design, usability, learnability, effectiveness, efficiency, satisfaction, visual, oral, aural, multimodal, auditory-icons, earcons, speech, text-to-speech, speech recognition, voice-instruction.

1 Introduction

Interfaces that offer interaction via more than one sense are highly demanded. The auditory system for example has been neglected in the development of user-interfaces [18]. Multimodality is needed for enhancing efficiency and effectiveness of user interface interaction, as well as it is needed for building more satisfactory and friendly interfaces. The aim of this study was to investigate the use of multimodal interaction metaphors (visual, oral, and aural) for improving learnability or the ability to complete tasks from first time use [13] in design environments. The investigation was performed by testing efficiency, effectiveness, and satisfaction of the proposed solutions. In order to perform this investigation, three design toolkits (TVOID, OFVOID, and MMID) offering visual, oral and aural interaction metaphors were built from scratch. An initial survey was used for taking evaluative views about the effectiveness and satisfaction of the interaction metaphors offered in the three toolkits. Then, a multi-group empirical study took place to complete this investigation. The following sections handle this scheme in more detail.

J. Jacko (Ed.): Human-Computer Interaction, Part III, HCII 2007, LNCS 4552, pp. 196–205, 2007.

2 Previous Work

The literature relevant to visual interface design has revealed the existence of usability drawbacks in existing interfaces of the visual-only design environments like cT™ [5], LAD [3], MICE [7], and CO-ED [6]. The heavily focus on conveying information through the visual channel when designing interfaces seems to make them "more and more visually crowded as the user's needs for communication with the computer increases" [17]. This causes the user to experience information overload by which important information may be missed [12]. The visual channel is not the only means that a human perceives information through. Studies on multimedia and its remarkable benefits for furthering the process of learning showed that multimedia can be used as an effective means for learning. A study by Rigas and Memery showed that multimedia helped users to learn more material than only text-and-graphics media, and assisted them in performing different tasks more successfully [22]. In order to solve complexity problems with the current visual user interfaces, Rigas et al [19] suggest that interfaces could be designed in away that visual metaphors communicate the information that 'needs' to be conveyed to the user and the auditory metaphors (earcons) communicate the other part of information (the interaction part) which is used to perform tasks like for example browsing e-mail data. The usability problems: miss-selection and interface intrusion into the task could be solved by employing auditory feedback, namely non-speech audio messages or earcons [1, 2]. The approach is based on using earcons to indicate the currently active graphical metaphor (menu, button, scrollbar, message-box, etc). It was found that the use of combinations of auditory icons, earcons, speech, and special sound effects helped users to make fewer mistakes when accomplishing their tasks, and in 'some cases' reduced the time taken to complete them [21]. The technology currently available for visually impaired users was also investigated, to explore the possibility of using it for enhancing usability of interfaces built for normal users. It was found that the ACCESS Project system by [14], in early 1990s, was pioneer in development of computer-based applications for blinds. It offered an audio-tactile environment for building such applications. Screen readers, which use synthesized speech, and Braille display were found to be the most used tools in blind users' applications. The Emacspeak [15] was the first project that introduced the conceptual-modeled screen reader that unlike traditional screen readers, which read screen contents only, it integrated spoken feedback with application contents. The Image Graphic Reader (IGR) by [16] outlined a procedure for reading charts and graphics for blind students. Haptic-mouse approach was employed after the IGR for reading information of charts and graphics to be, then, represented aurally [24]. A new interesting approach was introduced by [4] discovered that blind users could describe images and colors by listening to musical representations of visual images. Another study towards applying the same approach was carried out by [9] and [10] has given birth to a new invention called the vOICe. It differs from the one was introduced by Cronly-Dillon in that it implements conversion of highly complex video scenes, rather than still images, into sound, with the capability of determining colors. Previous studies have shown that the problems of usability in current user interfaces mainly occur because of lack of experience in designing multimodal interactive interfaces. The guidelines found in [8], [17], [20], [22], [11] have drawn a precise map for enhancing the way of conveying information to the user when using graphical widgets.

3 Experimental Toolkits

Three experimental toolkits were built from scratch using Microsoft Visual C# .NET. Each one of them provided a different set of interaction metaphors for doing the same functionality. These toolkits were developed to find out how learnable each of the implemented interaction metaphors can be, and which toolkit would be the most learnable. The following subsections handle these toolkits in more detail.

3.1 TVOID

TVOID or Typical Visual-Only Interface Designer imitates the style of interaction implemented in most of the existing interface-design systems like Microsoft Visual C# and Java NetBeans IDE. It interacts with the user visually-only with neglect to the other senses like the auditory system. In addition, it interacts with the user via six areas in its main interface, like most of the similar existing systems do. These areas are: menus, toolbar, toolbox, workplace (or drawing-area), properties-table, and status-bar. Figure 1.A shows a screenshot of TVOID.

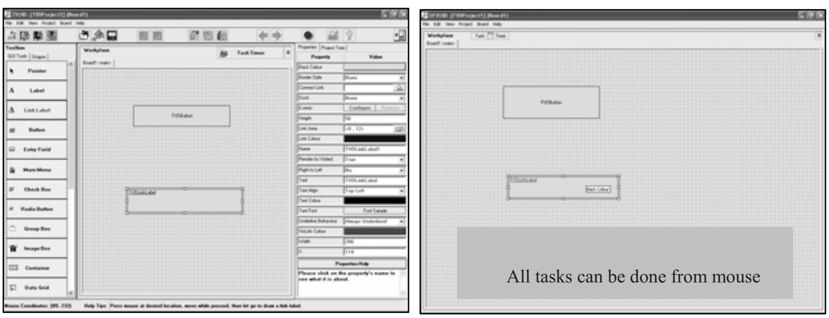

Fig. 1. Screenshots of the Visual-Only Interface Design Toolkits

3.2 OFVOID

OFVOID or On-the-Fly Visual-Only Interface Designer allows the user to do all design-tasks and the other tasks (menu/toolbar tasks) from the workplace area. There is no need for the mouse to leave the workplace area to do any job. The environment is facilitated with a number of time-saving features like, for example, selecting tools while drawing by scrolling over the form (board) being designed. The environment's main interface consists of two parts: menus and workplace. The menus part was added to the environment to show the users the hot-keys required to do the menu/toolbar functions. During the experiments, the users were allowed to use them. Figure 1.B shows a display of this environment.

3.3 MMID

MMID or Multi-Modal Interface Designer provides a combination of visual, vocal and aural interaction. This environment employs speech-recognition and text-to-speech for using it. It also allows the user to interact with the whole environment from the workplace area as the OFVOID environment. However, most of this interaction is through the use of voice-instruction and spoken messages. It adds another modality for interaction with less focus on visual interaction. Figure 2 shows a screenshot of MMID.

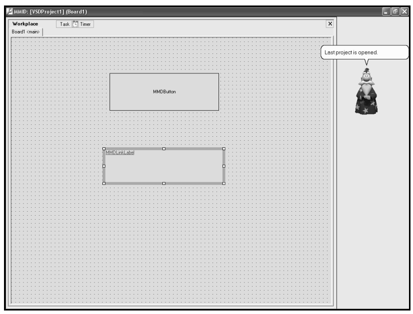

Fig. 2. Screenshot of the Multi-Modal Interface Designer (MMID)

4 Initial Survey

The initial survey aimed at taking users views about effectiveness and satisfaction of the interaction metaphors: auditory icons, earcons, spoken messages, and vocal commands, as solutions to solve some of the common usability problems. Thirty nine users participated in this survey. Their views were found valuable as they gave good impression about using sound (speech and non-speech) for conveying information to users who have no visual imparity. Further more, the initial survey showed to some extent how effective and satisfactory the use of voice-instruction for designing interfaces can be. Figure 3 shows the percentages of users who thought that the solutions implemented for solving the slipping-over buttons and hitting unwanted menu-items usability problems were effective. It also shows the participant’s views about effectiveness of visual and spoken help functionality-tags, and the feature implemented for conveying the “current-active-tool” piece of information.

Figure 4 shows that voice-instruction was not much appreciated for tool-selection and drawing. Nonetheless, it can be noticed that the users’ views had begun to change as the demonstration went on, and became more optimistic during setting properties by voice.

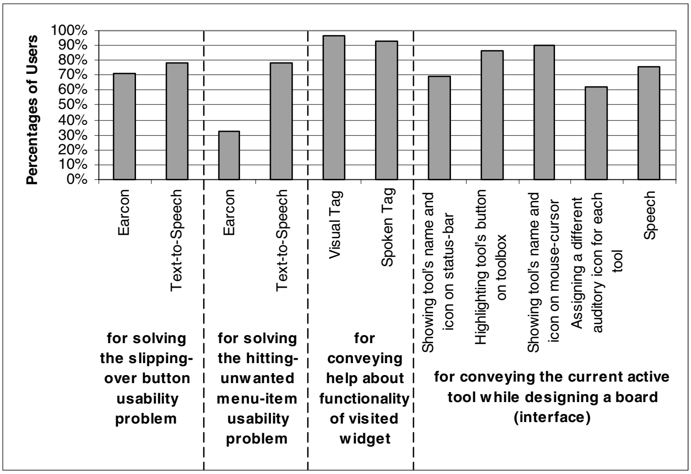

Fig. 3. Percentages of users who thought that the proposed solutions were effective for solving the shown usability problems

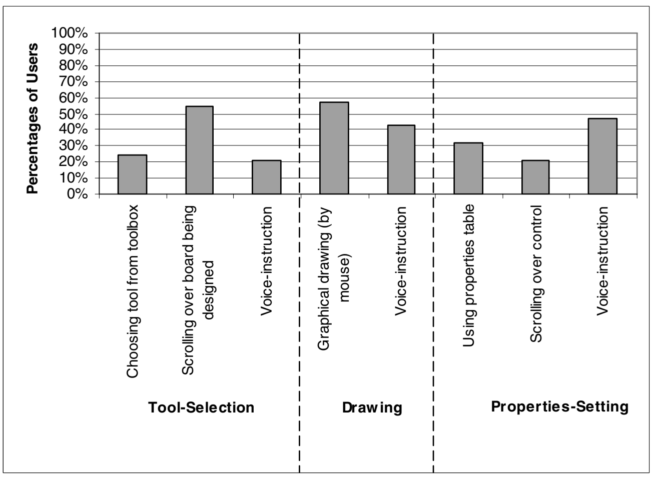

Fig. 4. Percentages of users who liked the demonstrated interaction metaphors for selecting tools, drawing, and setting properties

5 Empirical Multi-group Study

This study aimed at assessing the usability of each of the three design-toolkits to explore the most learnable one in regard to effectiveness, efficiency, and satisfaction. This assessment was fulfilled by testing these environments empirically by three independent groups of users. Each group consisted of 15 users. All the groups were asked to accomplish the same tasks (10 tasks). Each one of the users had attended a 10-minute video training session about the environment he/she was testing before doing the requested tasks. Effectiveness was measured by calculating the percentage of tasks completed successfully by all users and the percentage of functions learned in absence of additional help. Efficiency was measured by timing function learning and task completion for each task in each environment, and counting the number of errors during accomplishment of each task. At end of sessions, the users were asked to give ratings for satisfaction with the tested interaction metaphors.

6 Discussion of Results

During the experiments, it was noticed that the users in Group A, who tried the typical visual-only design environment TVOID expected how to do most of the functions. This environment looked familiar to them because they had experience with similar environments. This experience made them rely primarily on their memory. It made them spend time on recalling how to do particular functions in similar existing systems to be able to do them using TVOID. Expectations of how to do particular functions were incorrect sometimes, which caused the users to find out how to do these functions. In this way, the users who tested TVOID did two things to learn functions: remembering or expecting, and exploring in case of incorrect expectation. This was not the case with the users who tested OFVOID and MMID (Groups B and C). Most of the features in these two environments were new to them, which made them, most of time, head directly to exploring.

Also the results showed that the users in Groups B and C did more mistakes than the users in Group A during accomplishment of the same tasks. However, the difference was found to be not significant at 0.05 ($F = 1.31$, $P = 0.29$). In addition, it was noticed that the users were learning gradually about the interaction metaphors they were using in these two environments every time a task was accomplished. Making mistakes made them more used and familiar to the use of time-saving features and vocal commands. This caused the need for additional help to lessen in these two environments (OFVOID and MMID). Comparing the number of functions learned in absence of additional help under the three environments showed that MMID was the most learnable, and OFVOID was more learnable than the typical visual-only environment TVOID. Figure 5 shows this result.

The users who tested MMID needed less help because most of the vocal commands in the environment were as they expected, recalling that these commands were in the form of simple one-to-three-English phrases. In addition, looking for commands in one categorized list, as a feature in MMID, saved the time for looking for them in different positions, which was the case in the two visual-only environments. The frequent scrolling for selecting a particular tool to be used for drawing or a particular property to be set, in OFVOID, allowed the users to see and learn other tools and other properties on the mouse cursor every time a tool or property was selected. This made the users more familiar to this environment and lessened the need for additional help. The results also showed that the three environments were not as effective as each other. Gathering all vocal commands in one location (e.g. one list) in MMID helped the users in Group C to learn more functions than their counterparts in Groups B and C, who looked for commands in different locations. Also, the use of one interaction metaphor (voice-instruction) in MMID saved the time for the users to decide whether to use menus or toolbar,

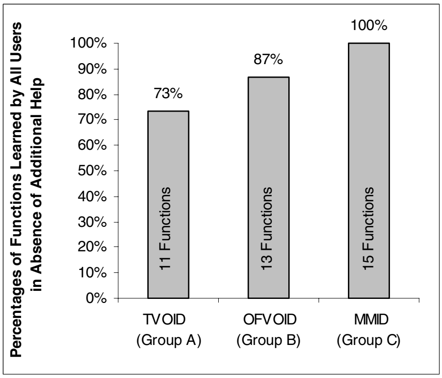

Fig. 5. Percentages calculated for the functions learned in absence of additional help by the users in Groups A (TVOID), B (OFVOID), and C (MMID)

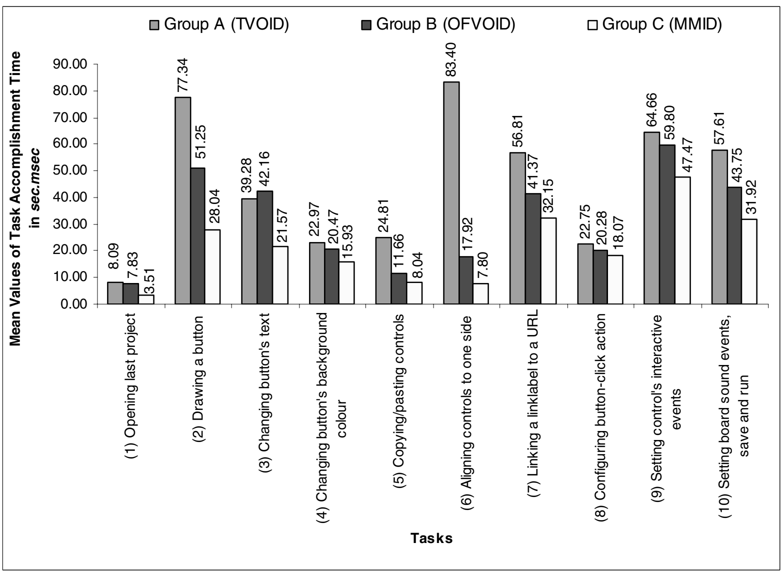

Fig. 6. Mean values of time taken for accomplishing 10 tasks for the first time using TVOID (Group A), OFVOID (Group B) and MMID (Group C)

toolbox or tool-list (by scrolling over board), properties table or properties-list (by scrolling over control), etc which were implemented in the two visual-only environments. The multimodal environment was more efficient in terms of shortening accomplishment time than the two visual-only environments. Figure 6 demonstrates this result. The use of vocal commands saved the time spent in moving the mouse from one position to another and scrolling menus and lists to reach commands in the two visual-only environments. Figure 7 shows the percentages of tasks completed successfully under each of the three environments. The difference between the three

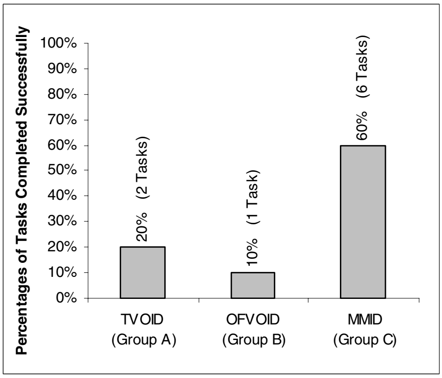

Fig. 7. Percentages of tasks completed successfully in TVOID (Group A), OFVOID (Group B), and MMID (Group C)

environments in this regard was found to be significant (F = 3.80, P = 0.04). The comments that were taken from the users in Group B during testing OFVOID showed that most of them expressed that the saving-time features were the most features they liked. On the other hand, the users who tested MMID (Group C) showed excitement toward designing by voice.

7 Conclusion and Future Research

The aim of this study was to investigate the use of multimodal interaction metaphors (visual, oral, and aural) for improving learnability of interface design environments. The paper presented a quick summary of relevant work. Then, it introduced a usability comparative study between multimodal interaction and visual-only interaction. The investigation started with a preliminary survey that aimed at taking users' views about the effectiveness and satisfaction of the proposed solutions for solving some of the common usability problems. Then an empirical multi-group study was introduced. This study aimed at testing the learnability parameters: effectiveness, efficiency, and satisfaction of a number of interaction metaphors offered by three design toolkits (TVOID, OFVOID, and MMID) built especially for the study. The paper, then, presented the results of these experiments and discussed them.

The study recommends designing design-environments bearing in mind interaction with the least possible mouse movement. A new effectively usable design environment is needed to save designers' time and effort and make them work satisfactorily. The use of time-saving interaction metaphors like instant scrollable lists, short-menus, and vocal commands can save much time during task-accomplishing and facilitate learning of new functions. The interactive mouse cursor was found to be a very good information conveyer as it can be used for showing the current active tool, the mouse coordinates, and the current active property to be set. This solution can extremely save the time for looking for this information in different places (i.e. the toolbox, the status-bar and the properties-table). However, the use of voice-instruction for selecting tools, drawing and setting properties was found to be more efficient and effective.

The empirical work covered in this paper investigated the usability of each of the three design environments from one angle, which is learnability or the ability to accomplish tasks from first time use. Further experiments will take place for testing Experienced User Performance (EUP) of task-completion in the three environments.

References

1. Beaudouin-Lafon, M., Conversy, S.: Auditory illusions for audio feedback. In: ACM CHI'96, Vancouver, Canada (1996)
2. Brewster, S., Clarke, C.V.: The design and evaluation of a sonically-enhanced tool palette. In: ICAD'97, Xerox PARC, USA (1997)
3. Chang, C., Chen, G., Liu, B., Ou, K.: A language for developing collaborative learning activities on World Wide Web. In: Proceedings of 20th International Computer Software and Applications Conference, COMPSAC '96 (1996)
4. Cronly-Dillon, J., Persaud, K.C., Blore, R.: Blind subjects construct conscious mental images of visual scenes encoded in musical form. In: The Royal Society Of London Series B-Biological Sciences, London (2000)
5. Cvetkovic, S.R., Seebold, R.J.A., Bateson, K.N., Okretic, V.K.: CAL programs developed in advanced programming environments for teaching electrical engineering, Education. IEEE Transactions on 37, 221–227 (1994)
6. Finkelstein, J., Nambu, S., Khare, R., Gupta, D.: CO-ED: a development platform for interactive patient education. In: Proceedings International Conference on Computers in Education, 2002 (2002)
7. Guercio, A., Arndt, T., Chang, S.-K.: A visual editor for multimedia application development. In: Proceedings. 22nd International Conference on Distributed Computing Systems Workshops, 2002 (2002)
8. Lumsden, J., Brewster, S., Crease, M., Gray, P.D.: Guidelines for Audio-Enhancement of Graphical User Interface Widgets. In: HCI'2002, London (2002)
9. Meijer, P.: Seeing with sound for the blind: Is it vision? In: Tucson 2002, Tucson, Arizona (2002)
10. Meijer, P.: Vision Technology for the Totally Blind, vol. 2004. Peter Meijer (2004)
11. Oakley, I., McGee, M.R., Brewster, S., Gray, P.D.: Putting the feel in look and feel. In: ACM CHI 2000, The Hague, NL (2000)
12. Oakley, I., Adams, A., Brewster, S., Gray, P.D.: Guidelines for the design of haptic widgets. In: BCS HCI 2002, London, UK (2002)
13. Payne, S.J., Green, T.R.G.: Task-Action Grammars: A Model of the Mental Representation of Task Languages. Human-Computer Interaction 2, 93–133 (1986)
14. Petrie, H., Morley, S., McNally, P., Graziani, P.: Authoring hypermedia systems for blind people. In: IEE Colloquium on Authoring and Application of Hypermedia-Based User-Interfaces, 1995 (1995)
15. Raman, T.V.: Emacspeak-an audio desktop. In: Proceeding IEEE Compcon '97 (1997)
16. Redeke, I.: Image and Graphic Reader. In: Proceedings 2001 International Conference on Image Processing, 2001 (2001)
17. Rigas, D.: Guidelines for Auditory Interface Design: An Empirical Investigation. In: Department of Computer Studies, University of Loughborough, Loughborough, p. 292 (1996)

18. Rigas, D., Memery, D., Yu, H.: Experiments in using structured musical sound, synthesised speech and environmental stimuli to communicate information: is there a case for integration and synergy? In: Proceedings of 2001 International Symposium on Intelligent Multimedia, Video and Speech Processing, 2001 (2001)
19. Rigas, D., Yu, H., Klearhou, K., Mistry, S.: Designing Information Systems with Audio-Visual Synergy: Empirical Results of Browsing E-Mail Data. In: Panhellenic Conference on Human-Computer Interaction. Advances on Human-Computer Interaction, Patras, Greece, 2001 (2001)
20. Rigas, D., Yu, H., Memery, D., Howden, D.: Combining Speech with Sound to Communicate Information in a Multimedia Stock Control System. In: 9th International Conference on Human-Computer Interaction: Usability Evaluation and Interface Design, New Orleans, Luisiana, USA, 2001 (2001)
21. Rigas, D., Hopwood, D., Yu, H.: The Role of Multimedia in Interfaces for On-Line Learning. In: 9th Panhellenic Conference on Informatics (PCI'2003), Thessaloniki, Greece, 2003 (2003)
22. Rigas, D., Memery, D.: Multimedia e-mail data browsing: the synergistic use of various forms of auditory stimuli. In: Proceedings. ITCC 2003. International Conference on Information Technology: Coding and Computing [Computers and Communications], 2003 (2003)
23. Rigas, D.I., Memery, D.: Multimedia e-mail data browsing: the synergistic use of various forms of auditory stimuli. In: Proceedings. ITCC 2003. International Conference on Information Technology: Coding and Computing [Computers and Communications], 2003 (2003)
24. Yu, W., Kangas, K., Brewster, S.: Web-based haptic applications for blind people to create virtual graphs. In: Proceedings. 11th Symposium on Haptic Interfaces for Virtual Environment and Teleoperator Systems, 2003. HAPTICS 2003 (2003)